Dynamic Pricing with Monotonicity Constraint Under Unknown Parametric Demand Model

Anonymous Authors

Abstract

We consider a Continuum-Armed Bandit problem with an additional monotonicity constraint (or "markdown" constraint) on the actions selected. This problem faithfully models a natural revenue management problem, called "markdown pricing", where the objective is to adaptively reduce the price over a finite sales horizon to maximize expected revenues. Chen ([4]) and Jia et al ([8]) recently showed a tight $T^{3/4}$ regret bound over T rounds under *minimal* assumptions of unimodality and Lipschitzness in the reward function. This bound shows that markdown pricing is strictly harder than unconstrained dynamic pricing (i.e., without the monotonicity constraint), which only suffers regret $T^{2/3}$ under the same assumptions ([10]). However, in practice, demand functions are usually assumed to have certain functional forms (e.g. linear or exponential), rendering the demand learning easier and suggesting lower regret bounds. We introduce a concept, *markdown dimension*, that measures the complexity of any parametric family, and present optimal regret bounds that improve upon the previous $T^{3/4}$ bound under this framework.

1 Introduction

Dynamic pricing under unknown demand arises naturally for the sale of new products, where the demand function is not available in advance. The seller in this case has to learn the demand function over time and faces a learning-vs-earning trade-off. This problem is therefore usually formulated as a Multi-Armed Bandit (MAB) problem. Although bandit problems have been well understood *theoretically*, in practice, however, we rarely see retailers implement such policies. This is a largely because some practical constraints are often overlooked by those policies. For example, price increases can potentially create a manipulative image of the retailer and negatively impact their ratings. For example, for online menu prices, "on average, a 1% price increase leads to 3-5% decrease in online ratings" ([12]). Therefore, retailers may sometimes implicitly face a monotonicity constraint (which we call "markdown constraint"), which requires that the prices selected be non-increasing. A pricing policy that satisfies such a constraint is usually referred to as *markdown pricing* policy.

Thus motivated, in this work, we consider the markdown pricing problem with unknown demand, under various assumptions. Although *unconstrained* dynamic pricing under unknown demand has been extensively studied, little is known about *markdown* pricing under unknown demand. Recently, [4] and [8] independently showed that unimodality and Lipschitzness in *revenue* function (defined as the price times the mean demand) are necessary to achieve sublinear regret. In this setting, any reasonable policy must reduce the price at a moderate rate and stop only when there is sufficient evidence for *overshooting* the optimal price. By selecting suitable parameters, [4, 8] showed that the above policy is indeed optimal, achieving a tight $T^{3/4}$ regret bound.

However, in practice, it is usually assumed that the demand function has certain parametric forms, such as linear, exponential or logit form, allowing faster demand learning and suggesting lower regret bounds. This motivates our first question:

Q1) Can we strengthen the $T^{3/4}$ regret bound for markdown pricing in this setting?

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Noticeably, the tight $T^{3/4}$ regret bound in [4, 8] this bound is asymptotically higher than $T^{2/3}$, the known regret bound for *unconstrained* pricing ([10]), highlighting the extra complexity caused by monotonicity constraint. Thus, we naturally arrive at our second question.

Q2) Is markdown pricing still harder than unconstrained pricing under parametric assumptions?

Or, more precisely, can we still show a *separation* between markdown and unconstrained pricing, under various parametric assumptions? While one may answer these two questions for specific parametric families, it is tempting to find a unified framework that captures the hardness of demand learning for a given parametric family. This motivates the following question.

Q3) Can we find a general framework to unify the regret bounds for different *categories* of families, rather than specific results for specific families?

In this work, we propose such a framework by introducing a concept called *markdown dimension*. We provide efficient markdown policies for each markdown dimension, which we also show to be *best* possible, thereby completely settling the problem of markdown pricing under unknown demand.

1.1 Our Contributions.

In this work, we make the following contributions.

1. **New Complexity Measure for Demand Learning:** We introduce a new concept called *markdown dimension*, which captures the complexity of performing markdown pricing on a family, answering Q3). Within this framework, we provide a complete settlement of the problem as specified below.
2. **Markdown Policies with Theoretical Guarantees:** For each finite markdown dimension $d \geq 0$, we present a efficient markdown pricing policy. Our policies proceed in *phases*, in which the seller learns the demand by selecting prices at a suitable spacing to estimate the true parameter, and then makes conservative decisions. We show that for $d = 0$ and $d \geq 1$, our policies achieve regret $O(\log^2 T)$ and $\tilde{O}(T^{\frac{d}{d+1}})$ respectively, settling Q1).
3. **Tight Minimax Lower Bounds:** We complement our upper bounds with a matching lower bound for each markdown dimension d. More precisely, we show that $\Omega(\log^2 T)$ regret is tight for $d = 0$, which *separates* it from the known $O(\log T)$ regret bound without this monotonicity constraint (see [3]). For finite $d \geq 1$, we show an $\Omega(T^{d/(d+1)})$ lower bound, which not only matches our upper bound but is also asymptoticly higher than the tight $\tilde{\Theta}(T^{1/2})$ upper bound (see [3, 9]) without markdown constraint, settling Q2).
4. **Impact of Smoothness:** We go further to refine our bounds and investigate the impact of smoothness of reward function around the optimal price. We consider a *sensitivity* parameter s, which essentially says that an ε distance away from the optimal price incurs $O(\varepsilon^s)$ regret, and can be verified to be at least 2 assuming the demand function is twice-differentiable. For both finite and infinite d, we extend our upper bounds to incorporate s. Moreover, for $d = \infty$, our tight $T^{\frac{2s+1}{3s+1}}$ regret bound is asymptoticly higher than that for unconstrained pricing, whose optimal regret is known to be $T^{\frac{s+1}{2s+1}}$ ([2]).

We highlight our results for $s = 2$ in red in Table 1, where Θ denotes tight upper and lower bounds.

Markdown Dimension	Markdown Pricing	Unconstr. Pricing
$d = 0$	$\Theta(\log^2 T)$	$\Theta(\log T)$
$1 \leq d < \infty$	$\tilde{\Theta}(T^{d/(d+1)})$	$\tilde{\Theta}(\sqrt{T})$
$d = \infty$	$\tilde{\Theta}(T^{3/4})$ [4, 8]	$\tilde{\Theta}(\sqrt{T})$ [3]

Table 1: Regret bounds for $s = 2$.

1.2 Related Work

In the *Multi-Armed Bandit* (MAB) problem, the player is offered a finite set of arms, with each arm providing a random revenue from an unknown probability distribution specific to that arm. The

objective of the player is to maximize the total revenue earned by pulling a sequence of arms (e.g. [11]). Our pricing problem generalizes this framework by using an infinite action space $[0, 1]$ with each price p corresponding to an action whose revenue is drawn from an unknown distribution with mean $R(p)$. In the *Lipschitz Bandit* problem (see, e.g., [1]), it is assumed that each $x \in [0, 1]$ corresponds to an arm with mean reward $\mu(x)$, and μ satisfies the Lipschitz condition, i.e. $|\mu(x) - \mu(y)| \leq L|x - y|$ for some constant $L > 0$. For for one-dimensional Lipschitz Bandits, there is a known $\tilde{\Theta}(T^{2/3})$ regret bound ([10]).

Recently there is an emerging line of work on bandits with monotonicity constraint. [4] and [8] recently independently considered the markdown pricing problem under unknown demand function and proved a tight $T^{3/4}$ regret bound under the minimal assumptions – Lipschitzness and unimodality on the revenue functions. Motivated by fairness constraints, [7] and [14] considered a general online convex optimization problem where the action sequence is required to be monotone.

Other requirements on the arm sequence motivated by practical problems are also considered in the literature. For example, [5, 13] and [15] are motivated by a similar concern in dynamic pricing: customers are hostile to frequent price changes. They propose algorithms that have a limited number of switches and study the impact on the regret. Motivated by Phase I clinical trials, [6] study the best arm identification problem when the reward of the arms are monotone and the goal is to identify the arm closest to a threshold.

2 Model and Assumptions

We begin by formally stating our model. In this work we assume an unlimited supply of a single product. Given a discrete time horizon of T rounds, in each round t, the policy (representing the "seller") selects a price p_t (the particular interval $[0, 1]$ is without loss of generality, by scaling). The demand D_t in this round is then independently drawn from a fixed distribution with *unknown* mean $D(p_t)$, and the policy receives revenue (or *reward*, which we will use interchangeably) p_t for each unit sold, and hence a total of $p_t \cdot D_t$ revenue in this round. The only constraint the policy must satisfy is the *markdown* constraint: $p_1 \geq \cdots \geq p_T$ almost surely.

The function $D(p)$ which maps each price p to the mean demand at this price is known as the *demand function*. For any policy π and demand function $D(\cdot)$, we use $r(\pi, D)$ to denote the expected total reward of π under D. Rather than evaluating policies directly in terms of $r(\pi, D)$, it is more informative (and ubiquitous in the literature on MAB) to measure performance using the notion of *regret* with respect to a certain idealized benchmark. Specifically, since we assumed unlimited supply, when the true reward function is known, the seller simply always selects a revenue-maximizing price $p^*_D = \arg\max_{p\in[0,1]} p \cdot D(p)$ at each round, and we denote this maximal reward rate to be $r^*_D = \max_{p\in[0,1]} p \cdot D(p)$. The regret of a policy is then defined with respect to this quantity, and we seek to bound the *worst-case* value over a given family of demand functions.

Definition 1 (Regret)**.** For any policy π and demand function D, define the *regret* of policy π under D to be $\mathrm{Reg}(\pi, D) := r^*_D \cdot T - r(\pi, D)$. For any given family $\mathcal{F}$ of demand functions, the *worst-case regret* (or simply *regret*) of policy π for family $\mathcal{F}$ is $\mathrm{Reg}(\pi, \mathcal{F}) := \sup_{D\in\mathcal{F}} \mathrm{Reg}(\pi, D)$.

2.1 Basic Assumptions

Now we state the common assumptions that all of our results rely on. A demand function $D(\cdot)$ is naturally associated with a *revenue function* (or, *reward function*) $R(p) = p \cdot D(p)$. Sometimes it will be convenient to work directly with the revenue functions, in which cases we write $r(\pi, R)$ as the regret of under reward function is R.

Definition 2 (Optimal Price Mapping)**.** Let $\mathcal{F}$ be a set of functions defined on some set $S \subseteq \mathbb{R}$. For any function $R : S \to \mathbb{R}$, let $M(R)$ be the subset of its global maxima on S. The *optimal price mapping* $p^* : \mathcal{F} \to S$ is defined as $p^*(R) = \inf M(R)$.

By elementary topology, if the domain S is compact, then $M(R)$ is also compact, so the infimum of $M(R)$ can be attained, and hence $p^*(R)$ is also a global maximum of R. We first introduce a standard assumption (see e.g. [3]), which assumes the derivative of R vanishes at $p^*(R)$.

Assumption 1 (Vanishing Derivative)**.** We assume that every reward function R is differentiable on its domain, and moreover, $R'(p^*(R)) = 0$.

In particular, this assumption holds if the reward function attains global its optimum in the interior of the domain. Under Assumption 1, by Taylor expansion on R around p^*, for small ε we have

$$|R(p^*) - R(p^* + \varepsilon)| = \frac{1}{2}R''(p^*) \cdot \varepsilon^2 + o(\varepsilon^2).$$

Thus, if a policy overshoots the optimal price by ε, then only an $O(\varepsilon^2)$ loss is incurred in each round. We next introduce a *distributional* assumption.

Definition 3 (Subgaussian Random Variable)**.** The *subgaussian norm* of a random variable X is

$$\|X\|_{\psi_2} := \inf\{c > 0 : \mathbb{E}[e^{X^2/c^2}] \leq 2\},$$

and X is said to be *subgaussian* if $\|X\|_{\psi_2} < \infty$.

Assumption 2 (Subgaussian noise)**.** There exists a constant $C_{sg} > 0$ such that under any true demand function and any price p, the random demand X at price p satisfies $\|X\|_{\psi_2} \leq C_{sg}$.

For example, this assumption is satisfied when the demand distributions are all Bernoulli, or when they are normal distributions with bounded variances.

2.2 Identifiability and Robust Parametrization

Our goal in the rest of this section is to develop a novel concept, *markdown dimension*, that characterizes the complexity of a family of reward functions. In this subsection, we explain the high level ideas behind our definition.

Intuitively, the exploration-exploitation trade-off for markdown pricing becomes harder to manage as the given family becomes more complex. Consider, for example, linear demand functions. If each function takes the form $D(p; c) = 1 - cp$ where only $c \in (0, 1)$ is unknown and $p \in [0, 1]$, then the seller simply needs to estimate the (negative) slope c by sampling sufficiently many times at $p = 1$.

In contrast, if each function takes the form $D(p; a, b) = a - bp$ where *both* parameters a, b are unknown, then sampling at one price would *not* suffice. Rather, one needs to select (at least) two distinct prices to estimate a, b, thereby facing the following dilemma. Suppose the two prices $p < p'$ selected are far apart. Then, p may be far away from the optimal price p^* since p^* may be close to p', resulting in high regret. Otherwise, when those prices are close by, the demand learning requires a high volume of samples, which potentially also leads to a high regret.

Thus we reach a natural question: can we introduce a complexity index to measure the difficulty of performing markdown pricing on a given family, and then provide tight regret bounds in terms of this complexity index? In this work, we propose a complexity index, called *markdown dimension*, and provide nearly-optimal regret bounds in terms of the markdown dimension of the given family of demand functions. The formal definition relies on other two concepts, the *identifiability* of a family, and the *robustness* of a parametrization, which we introduce in the next two subsections.

Identifiability Our notion of identifiability generalizes a key property of single-variable polynomials, that every degree-d polynomial can be uniquely determined by its values at *any* $(d + 1)$ points. To present the formal definition, we first introduce a mapping which, for a fixed subset of prices, assigns each demand function a *profile* based on its values at those prices.

Definition 4 (Profile Mapping)**.** Consider a set $\mathcal{F}$ of real-valued functions defined on $A \subseteq \mathbb{R}$. For any fixed $\mathbf{p} = (p_0, p_1, ..., p_d) \in A^{d+1}$, the *profile-mapping* with respect to $\mathbf{p}$ is defined as

$$\begin{aligned} \Phi_{\mathbf{p}} : \mathcal{F} &\to \mathbb{R}^d, \\ D &\mapsto (D(p_0), D(p_1), ..., D(p_d)). \end{aligned}$$

We may subsequently call $\Phi_{\mathbf{p}}(D)$ the *profile* of function D with respect to $\mathbf{p}$. In words, the above says that every function in $\mathcal{F}$ be assigned a unique profile at any $(d + 1)$ distinct points ("prices").

Definition 5 (Identifiability)**.** The family $\mathcal{F}$ is *d-identifiable*, if for any *distinct* $p_0, p_1 ..., p_d \in S$, the profile mapping $\Phi_{p_0,...,p_d}$ is injective, i.e. distinct functions in $\mathcal{F}$ are mapped to distinct profiles.

In particular, if a family is d-identifiable, then for any distinct $p_0, p_1 ..., p_d$ the inverse profile-mapping $\Phi_{\mathbf{p}}^{-1} : \mathcal{R}_p \to \mathcal{F}$ exists, where $\mathcal{R}_p$ is the range of the mapping $\Phi_{\mathbf{p}}$.

Robust Parametrization We first formally define a *parametrization*.

Definition 6 (Parametrization)**.** An *order-m parametrization* for a family $\mathcal{F}$ of functions is any one-to-one mapping from a compact set $\Theta \subseteq \mathbb{R}^m$ to $\mathcal{F}$. Moreover, each $\theta \in \Theta$ is called a *parameter*.

By abuse of notations, we may use $D(p;\theta)$ to denote the function $D(p)$ that parameter θ corresponds to. As a standard assumption (see e.g. [3]), we also assume that the parameter set Θ to be compact, which leads to many favorable properties.

Assumption 3 (Compact Domain)**.** The domain Θ of the parametrization is compact.

Under this assumption, the demand functions in $\mathcal{F}$ are bounded, and thus we may without loss of generality also scale the range (i.e. target space) of those functions to be $[0,1]$.

Assumption 4 (Smoothness)**.** The mapping $D : [0,1] \times \Theta \to \mathbb{R}$ is twice-differentiable and admits continuous second partial derivative. In particular, since $[0,1] \times \Theta$ is compact, under the above assumption, there exist constants $C^{(j)} > 0$ such that the j-th derivative satisfies $|D^{(j)}(p,\theta)| \leq C^{(j)}$ for any $(p,\theta) \in [0,1] \times \Theta$ and $j = 0,1,2$.

Recall that we previously defined the optimal price mapping p^* from $\mathcal{F}$ to the domain, $[0,1]$, of the reward functions. Now that we introduced a parametrization, by abuse of notation we may view the mapping p^* as being defined on $\Theta \subseteq \mathbb{R}^m$, which enables us to consider its Lipschitzness.

Assumption 5 (Lipschitz Optimal Price Mapping)**.** The optimal price mapping $p^* : \Theta \to [0,1]$ is C^*-Lipschitz for some constant $C^* > 0$.

In particular, by compactness of Θ, a sufficient condition for the above to hold is that p^* admits continuous first-order partial derivatives. This assumption is hence satisfied by most "smooth" demand functions. For example, take linear demand function $D(p;c) = 1 - cp$ where $p, c \in [0,1]$. Then, the optimal price is $p^*(c) = \min\{\frac{1}{2c}, 1\}$, which can be easily seen to be 1-Lipschitz.

The final ingredient for robust parametrization is motivated by the following robustness of the *natural parametrization* $D(p;\theta) = \sum_{j=0}^{d} \theta_j p^j$ for polynomials. Consider any distinct prices $p_0, p_1, ..., p_d$, and any $d+1$ real numbers $y_0, y_1, ..., y_d$ representing, for example, the mean reward at each p_i. We may then uniquely determine a degree-d polynomial by solving the linear equation

$$\begin{bmatrix} 1 & p_0 & p_0^2 \cdots & p_0^d \\ 1 & p_1 & p_1^2 \cdots & p_1^d \\ & & \vdots & \\ 1 & p_d & p_d^2 \cdots & p_d^d \end{bmatrix} \cdot \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_d \end{bmatrix}.$$

The matrix on the left-hand side is often referred to as the *Vandermonde* matrix, denoted $V_{\mathbf{p}} := V(p_0, ..., p_d)$. One can easily verify that when p_i's are distinct, $V_{\mathbf{p}}$ has non-zero determinant and is hence invertible, and hence

$$\theta = V_{\mathbf{p}}^{-1}\mathbf{y},$$

where $\mathbf{y} = (y_0, y_1, ..., y_d)$. Next consider the effect of a perturbation on $\mathbf{y}$, in terms of the following *separability* parameter.

Definition 7. For any $\mathbf{p} = (p_0, ..., p_d) \in \mathbb{R}^{d+1}$, define $h(\mathbf{p}) := \min_{i \neq j} |p_i - p_j|$.

To motivate the notion of robust parametrization, let us first consider a result that is specific to polynomial demand functions.

Proposition 1. There exist constants $C_1, C_2 > 0$ such that for any $\mathbf{p} \in \mathbb{R}^{d+1}$ with $0 < h(\mathbf{p}) \leq C_1$, any $\varepsilon \leq C_1$, and $\mathbf{y}, \hat{\mathbf{y}} \in \mathcal{R}_p$ with $\|\mathbf{y} - \hat{\mathbf{y}}\|_\infty \leq C_1$, it holds that

$$\|V_{\mathbf{p}}^{-1}\mathbf{y} - V_{\mathbf{p}}^{-1}\hat{\mathbf{y}}\|_\infty \leq C_2 \cdot \|\mathbf{y} - \hat{\mathbf{y}}\|_\infty \cdot h(\mathbf{p})^{-d}. \tag{1}$$

More concretely, let $D(p,\theta)$ be the underlying polynomial demand function, and $\mathbf{y} = V_p \cdot \theta$. Suppose $\hat{\mathbf{y}}$ corresponds to the empirical mean demands at prices $p_0, ..., p_d$. Then, one can estimate θ using the plug-in estimator $V_{\mathbf{p}}^{-1} \cdot \hat{\mathbf{y}}$. Proposition 1 then upper bounds the estimation error of this estimator, in terms of $h(\mathbf{p})$ and $\mathbf{y} - \hat{\mathbf{y}}$.

In order to achieve sublinear regret, the value $h = h(\mathbf{p})$ must be chosen to be $o(1)$ as T goes to infinity. Thus, the dependence on h crucially affects our regret bounds. Proposition 1 establishes a

nice property for polynomials, that the estimation error increases in the order of $(\frac{1}{h})^d$, as $h \to 0^+$. We introduce *robust parametrization* by generalizing this property beyond polynomials. Loosely, an order-d parametrization is *robust*, if it admits a similar error bound to (1).

Definition 8 (Robust Parametrization)**.** An order-d parametrization $\theta : \Theta \to \mathcal{F}$ is *robust*, if
(1) it satisfies Assumptions 3, 4 and 5, and
(2)there exist constants $C_1, C_2 > 0$ such that for any $\mathbf{p} \in \mathbb{R}^{d+1}$ with $0 < h(\mathbf{p}) \leq C_1$ and any $y, y' \in \mathcal{R}_p$ with $\|\mathbf{y} - \mathbf{y}'\|_\infty \leq C_1$, it holds that $\|\Phi_{\mathbf{p}}^{-1}(\mathbf{y}) - \Phi_{\mathbf{p}}^{-1}(\mathbf{y}')\|_\infty \leq C_2 \cdot \|\mathbf{y} - \mathbf{y}'\|_\infty \cdot h(\mathbf{p})^{-d}$.

2.3 Markdown Dimension

Now we are ready to define the markdown dimension.

Definition 9 (Markdown Dimension)**.** The *markdown dimension* (or simply *dimension*) for a family $\mathcal{F}$ of functions, denoted $d(\mathcal{F})$, is the minimum integer $d \geq 0$ such that $\mathcal{F}$ is (i) d-identifiable, and (ii) admits a robust order-d parametrization. If no finite d satisfies the above conditions, then $d(\mathcal{F}) = \infty$.

For example, under mild assumptions, any family of degree-d polynomials is d-dimensional under the natural parametrization.

Proposition 2. Let $\mathcal{F} = \{\sum_{j=0}^{d} \theta_j x^j \mid \theta \in \Theta\}$ where Θ is compact. Then, $\mathcal{F}$ is d-dimensional.

We further illustrate our definition by considering the dimensions of some commonly used families. As the simplest family, one may verify that our definition of 0-dimensional family (under our assumptions) is equivalent to the *separable* family as defined in Section 4 of [3]. We provide more concrete examples below.

Proposition 3. The following families are 0-dimensional:
(1) single-parameter linear demand functions: $\mathcal{F}_0 = \{1 - ax : b \in [a_{min}, a_{max}]\}$,
(2) exponential demand functions: $\mathcal{F}_1 = \{e^{1-ax} : b \in [a_{min}, a_{max}]\}$,
(3) logit demand functions: $\mathcal{F}_2 = \{\frac{e^{1-ax}}{1+e^{1-ax}} : b \in [a_{min}, a_{max}]\}$.

Finally, we observe that by our definition, if a family of functions is not d-identifiable for any d, then it is infinite dimensional, as illustrated by the following example.

Proposition 4. Let $\mathcal{F}$ be the set of all 1-Lipschitz functions on $[0, 1]$, then $d(\mathcal{F}) = \infty$.

In this work, for each finite $d = 0, 1, 2, ...$, we will propose an efficient markdown pricing policy for dimension d families, which we also prove to be the best possible theoretically.

2.4 Sensitivity

Consider the Taylor expansion of a reward function $R(x)$ around an optimal price p^*:

$$R(p) = R(p^*) + 0 + \frac{1}{2!}R''(p^*)(p - p^*)^2 + \frac{1}{3!}R^{(3)}(p^*)(p - p^*)^3 + ...$$

Suppose the first nonzero derivative is $R^{(k)}(p^*)$. Then, the higher k, the less the revenue is *sensitive* to overshooting (i.e. $p < p^*$). This motivates us to introduce the following concept, *sensitivity*, that measures how fast the revenue function changes around the optimal price.

Definition 10 (Sensitivity)**.** A reward function is called *s-sensitive* if every function $R \in \mathcal{F}$ is $(s + 1)$-differentiable, with $R^{(1)}(p^*(R)) = ... = R^{(s-1)}(p^*(R)) = 0$ and $R^{(s)}(p^*(R)) < 0$. A family $\mathcal{F}$ of reward functions is called *s-sensitive* if
(a) every $R \in \mathcal{F}$ is s-sensitive,
(b) it admits a parametrization $R(x; \theta)$ satisfying Assumptions 3 to 5, and
(c) there is a constant $C_6 > 0$ such that $R^{(s)}(p^*(R)) \leq -C_6 < 0$ for any $R \in \mathcal{F}$.

For example, let $R(x; \theta) = \theta - |\frac{1}{2} - x|^s$ for $x \in [0, 1]$ and any $\theta \in \mathbb{R}$, then $\{R(x; \theta) : \theta \in [\frac{1}{2}, 1]\}$ is an s-sensitive family.

Note that by Taylor's Theorem, $|R(p^* + \varepsilon) - R(p^*)| \leq \frac{C_s}{s!}|\varepsilon|^s$. Consequently, if a policy overshoots or undershoots the optimal price by ε, the regret *per round* is only $O(\varepsilon^s)$, which is asymptoticly lower than the per-round regret $O(\varepsilon^2)$ without the sensitivity assumption. For unconstrained pricing

(or continuum bandits), [10] showed that a tight $T^{\frac{s+1}{2s+1}}$ regret for s-sensitive reward functions. In particular, for $s = 1$, the regret becomes $T^{2/3}$, strictly lower than the $T^{3/4}$ result. In this work we address a natural question: how does the regret bounds for markdown pricing change as s increases?

We give an overview of our policies and prove regret upper bounds for zero, finite (non-zero) and infinite dimensional families of demand functions. Moreover, we show that our policies achieves nearly optimal regret by providing lower bounds for each of these regimes.

3 Zero-Dimensional Family

We start with the simplest case, 0-dimensional demand functions. We propose a policy called *Cautious Myopic* which proceeds by *phases* and makes *conservative* decisions. As opposed to the *optimism* in the face of uncertainty in UCB type policies, our policy adopts *conservatism* in the face of uncertainty.

More precisely, we partition the time horizon so that the j-th phase consists of $t_j := \lceil 2^j \log T \rceil$ rounds (for simplicity we assume T is a multiple of $\lceil \log T \rceil$), and thus in total there are $K = O(\log T - \log\log T)$ phases. In each phase, the policy estimates the true parameter θ^* using the observations from the last phase, and builds a confidence interval around θ^*, which depends on the number of length of this phase and also the constant C_{sg} as defined in Assumption 2. Then, in the next phase, the policy selects the *largest* optimal price of any parameter θ in the confidence interval. We write $t^{(j)} := \sum_{k=0}^{j} t_k$ and for convenience $t^{(0)} = 0$, and formally state this policy in Algorithm 1.

Algorithm 1 Cautious Myopic Policy.

Input: a family $\mathcal{F}$ of demand functions and time horizon T.
$p_1 \leftarrow 1$ % Initialization
for $j = 1, ..., K$ **do**
  for $t = t^{(j-1)} + 1, ..., t^{(j-1)} + t_j$ **do**
    $x_t \leftarrow p_j$ % Select p_j for t_j times in phase j
    Observe realized demand D_t
  end for
  $\bar{d}_j = \frac{1}{t_j} \sum_{\tau=1}^{t_j} D_{t^{(j-1)}+\tau}$ % Empirical mean demand in phase j
  $\hat{\theta}_j \leftarrow \Phi_{p_j}^{-1}(\bar{d}_j)$ % Estimate parameter
  $w_j \leftarrow 2C_{sg}\sqrt{\frac{\log T}{t_j}}$ % Width of the confidence interval
  $p_{j+1} \leftarrow \max\{p^*(\theta) : |\theta - \hat{\theta}_j| \leq w_j\}$ % Conservative estimation of the optimal price
end for

Theorem 1 (Zero-dimensional Upper Bound)**.** Let $\mathcal{F}$ be any 0-dimensional, s-sensitive family of demand functions. Then the Cautious Myopic (CM) Policy has regret

$$\mathrm{Reg}(\mathrm{CM}, \mathcal{F}) = \begin{cases} O(\log^2 T), & \text{if } s = 2, \\ O(\log T), & \text{if } s > 2. \end{cases}$$

As we discussed earlier, many simple families of demand functions such as single-parameter linear or exponential demand functions satisfy $s = 2$, thereby having regret $O(\log^2 T)$.

It is worth noting that this bound is asymptotically higher than the $O(\log T)$ upper bound for unconstrained pricing ([3]). Intuitively, this is because the CM policy purposely keeps a distance from the estimated optimal price. But can we achieve $o(\log^2 T)$ regret by taking more risk or being more conservative? We answer this question negatively by showing that CM is indeed optimal, providing a *separation* unconstrainted and markdown pricing for 0-dimensional demand families.

Theorem 2 (Zero-Dimensional Lower Bound)**.** There exists a 0-dimensional 2-sensitive family $\mathcal{F}$ such that for any policy π, we have $\mathrm{Reg}(\pi, \mathcal{F}) = \Omega(\log^2 T)$.

4 Finite-Dimensional Family

Now we consider finite, nonzero dimensional families. For dimension d, the learner needs to collect demand samples at $d+1$ distinct *sample prices*, as opposed to just one price when $d=0$. This, however, incurs extra regret, since the optimal price may lie *between* these sample prices.

Intuitively, a reasonable policy trades off between the overshooting risk and the learning rate. If the gap is large, the policy may learn the parameter efficiently, but there is potentially a higher regret due to overshooting, in case the true optimal price lies between the sample prices. On the other extreme, if the gap is small, there is less risk of overshooting but a slower rate of learning.

We introduce our Iterative Cautious Myopic (ICM) Policy (Algorithm 2) that strikes such balance nearly optimally, as we will soon see. The policy consists of m phases. In phase $j \in [m]$, the policy selects d *sample prices* evenly spaced at distance h, each for T_j times. It then computes an estimated the optimal price $\hat{p}_j$ along with a confidence interval $[L_j, U_j]$, based on the observed demands.

To determine the initial price p_{j+1} in the next phase, the policy considers three cases. Recall that the last price that the policy selects in phase j is $p_j - dh$. We say a *good* event occurs, if $p_j - dh > U_j$, in which case we set $p_{j+1} = U_j$. In the *dangerous* event, the current price $p_j - dh$ is within the confidence interval, and we may have already overshot the optimal price. Thus, the policy needs to behave conservatively to prevent from overshooting further, so it selects $p_{j+1} = p_j - dh$. Finally in the *overshooting* event, the current price is already lower than L_j, and hence with high probability the policy has overshot. In this case, it immediately exits the exploration phase (i.e. the outer for-loop) and enters the exploitation phase, wherein it selects the current price in all remaining rounds.

Algorithm 2 Iterative Cautious Myopic Policy.

Input: $\mathcal{F}, m, \{T_j\}_{j\in[m]}, T$
$p_1 \leftarrow 1, L_0 \leftarrow 0, U_0 \leftarrow 1$ Initialization
for $j = 1, 2, ...m$ **do**
  for $k = 0, 1, ..., d$ **do**
    Select price $p_j - kh$ for T_j times and observe $D_1, ..., D_{T_j}$
    $\bar{D}_k \leftarrow \frac{1}{T_j}\sum_{i=1}^{T_j} D_i$ % Mean demand at $p_j - kh$
  end for
  $\hat{\theta} \leftarrow \Phi^{-1}_{p_j,\ldots,p_j-dh}(\bar{D}_0, ..., \bar{D}_d)$ % Estimate Parameter
  $w_j \leftarrow 2h^{-d} \cdot C_2 \cdot C_{sg}\sqrt{\frac{d\log T}{T_j}}$ % Width of confidence interval
  $L_j \leftarrow \min\{p^*(\theta) : \|\theta - p^*(\hat{\theta})\|_2 \le w_j\}$ % Lower confidence bound
  $U_j \leftarrow \max\{p^*(\theta) : \|\theta - p^*(\hat{\theta})\|_2 \le w_j\}$ % Upper confidence bound
  if $U_j \le p_j - dh$ **then**
    $p_{j+1} \leftarrow U_j$ % Good event
  end if
  if $U_j > p_j - dh \ge L_j$ **then**
    $p_{j+1} \leftarrow p_j - dh$
  end if
  if $p_j - dh < L_j$ **then**
    Break % Overshooting event
  end if
end for
Select the current price in every future roundExploitation

Theorem 3 (Upper Bound for Finite $d \ge 1$)**.** For any $m = \tilde{O}(1)$, there exists suitable choice of $h > 0$ and $T_1 < ... < T_m$, such that the Iterative Cautious Myopic Policy Policy ICM $=$ ICM$(T_1, ..., T_m, h)$ achieves regret $\mathrm{Reg}(\mathrm{ICM}, \mathcal{F}) = \tilde{O}\left(T^{\rho(m,s,d)}\right)$ where

$$\rho(m,s,d) = \frac{1 + \left(1 + \frac{s}{2} + ... + (\frac{s}{2})^{m-1}\right) d}{\left(\frac{s}{2}\right)^m + \left(1 + \frac{s}{2} + ... + (\frac{s}{2})^{m-1}\right) \cdot (d+1)}.$$

In particular, for $s = 2$ and $m = \log T$, we have $\mathrm{Reg}\left(\mathrm{ICM}, \mathcal{F}\right) = \tilde{O}\left(T^{\frac{d}{d+1}}\right)$.

We complement the upper bound with an nearly tight lower bound, up to $\log T$ factors, stated below.

Theorem 4 (Lower Bound for Finite $d \geq 1$)**.** For any $d \geq 2$, there exists a d-dimensional family $\mathcal{F}$ of demand functions on $[0, 1]$ such that for any markdown policy π, we have $\mathrm{Reg}(\pi, \mathcal{F}) = \Omega(T^{\frac{d}{d+1}})$.

In our proof, for each $d \geq 1$ we construct a sub-family of $(d+1)$-degree decreasing polynomial demand functions – which is also d-dimensional – on which any policy suffers regret $\Omega(T^{\frac{d}{d+1}})$.

5 Infinite Dimensional Family

When $d = \infty$, it is more convenient to work with the reward function $R(x) := x \cdot D(x)$ instead of the demand function. In contrast to *unconstrained* pricing, there is no *markdown* policy that achieves $o(T)$ regret on the family of Lipschitz reward functions ([4, 8]), since the reward function may have multiple local maxima. Nonetheless, [4] and [8] showed that if in addition we assume the reward function to be unimodal (which is satisfied by many commonly used families), then a tight $\tilde{\Theta}(T^{3/4})$ regret is achievable. In this setting, the problem boils down to finding the unique local optimum p^* of the true revenue function. Specifically, the lower bound is established by considering reward functions where the reward rate may change *abruptly* at p^*. We next show that the regret bound can be improved if we assume that the reward functions change smoothly.

Algorithm 3 Uniform Elimination Policy ($\mathrm{UE}_{m,\Delta}$).

Input: $T, \Delta, m > 0$
Initialize: $L_{\max} \leftarrow 0$, $w \leftarrow 2C_{sg}\sqrt{\frac{\log T}{m}}$
for $j = 0, 1, 2, ..., \lceil \Delta^{-1} \rceil$ **do**
  $x_j \leftarrow 1 - j\Delta$
  Select price x_j for the next m rounds and observe rewards $Z_1^j, ..., Z_m^j$
  $\bar{\mu}_j \leftarrow \frac{1}{m}\sum_{i=1}^{m} Z_i^j$ % Compute mean rewards
  $[L_j, U_j] \leftarrow [\bar{\mu}_j - w, \bar{\mu} + w]$ Compute confidence interval for reward at current price
  if $L_j > L_{\max}$ **then**
    $L_{\max} \leftarrow L_j$ % Keep track of the highest L_j
  end if
  if $U_j < L_{\max}$Exploration phase ends **then**
    $h \leftarrow j$ % Define the *halting price*
    Break
  end if
end for
Select price x_h in all remaining rounds. % Exploitation phase

Theorem 5 (Upper Bound for Infinite-Dimensional Family)**.** Let $\mathcal{F}_s^U$ be the family of unimodal and s-sensitive reward functions. For any $s \geq 2$, there exists suitable choices of m and Δ such that $\mathrm{Reg}(\mathrm{UE}_{m,\Delta}, \mathcal{F}_s^U) = O(T^{\frac{2s+1}{3s+1}})$.

We complement the above theorem with a lower bound in terms of both s and T, that matches the upper bound in Theorem 5 for every $s \geq 2$.

Theorem 6 (Lower Bound for s-Sensitive Family)**.** For any $s \geq 2$, there is a family $\mathcal{F}$ of s-sensitive unimodal revenue curves such that any markdown policy π satisfies $\mathrm{Reg}(\pi, \mathcal{F}) = \Omega(T^{\frac{2s+1}{3s+1}})$.

References

[1] R. Agrawal. The continuum-armed bandit problem. *SIAM journal on control and optimization*, 33(6):1926–1951, 1995.

[2] P. Auer, R. Ortner, and C. Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *International Conference on Computational Learning Theory*, pages 454–468. Springer, 2007.

[3] J. Broder and P. Rusmevichientong. Dynamic pricing under a general parametric choice model. *Operations Research*, 60(4):965–980, 2012.

[4] N. Chen. Multi-armed bandit requiring monotone arm sequences. *Advances in Neural Information Processing Systems*, 34, 2021.

[5] W. C. Cheung, D. Simchi-Levi, and H. Wang. Dynamic pricing and demand learning with limited price experimentation. *Operations Research*, 65(6):1722–1731, 2017.

[6] A. Garivier, P. Ménard, L. Rossi, and P. Menard. Thresholding bandit for dose-ranging: The impact of monotonicity. *arXiv preprint arXiv:1711.04454*, 2017.

[7] S. Gupta and V. Kamble. Individual fairness in hindsight. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 805–806, 2019.

[8] S. Jia, A. Li, and R. Ravi. Markdown pricing under unknown demand. *Available at SSRN 3861379*, 2021.

[9] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 594–605. IEEE, 2003.

[10] R. D. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704, 2005.

[11] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[12] M. Luca and O. Reshef. The effect of price on firm reputation. *Management Science*, 2021.

[13] G. Perakis and D. Singhvi. Dynamic pricing with unknown non-parametric demand and limited price changes. *Available at SSRN 3336949*, 2019.

[14] J. Salem, S. Gupta, and V. Kamble. Taming wild price fluctuations: Monotone stochastic convex optimization with bandit feedback. *arXiv preprint arXiv:2103.09287*, 2021.

[15] D. Simchi-Levi and Y. Xu. Phase transitions and cyclic phenomena in bandits with switching constraints. *Advances in Neural Information Processing Systems*, 32, 2019.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] See supplementary file
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]